昆 虫 学 报 *Acta Entomologica Sinica*，April 2006，49(2)：172 – 178　　ISSN 0454-6296

家蚕组织蛋白酶 D 基因的克隆、序列分析及其表达谱研究

杨远萍，刘　春，王子龙，王根洪，夏庆友*

（西南大学蚕学与生物技术学院，农业部蚕桑学重点实验室，重庆　400716）

摘要：**组织蛋白酶** D（cathepsin D，CtD）**是溶酶体内天冬氨酸内切蛋白酶，参与机体多种生理病理过程，尤其在昆虫的发育变态过程中起着重要作用。利用** NCBI **上登录的组织蛋白酶** D **基因核酸序列和家蚕** *Bombyx mori* **表达序列标签**（expressed sequence tags，EST）**数据库，进行电子克隆获得家蚕组织蛋白酶** D（*Bm*CtD）**基因的全长** cDNA（DQ010007）。**该** cDNA **大小为** 1 543 bp，**其中** ORF **长** 1 152 bp，**同源性分析表明** *Bm*CtD **与其他物种的** CtD **具有较高的相似性。** *Bm*CtD **的** mRNA **存在选择性拼接，另外一种** mRNA **形式命名为** *Bm*CtD Ⅰ。RT-PCR **实验表明该基因在本实验所调查的家蚕不同发育时期和组织中都有表达。**

关键词：**家蚕；组织蛋白酶** D；**克隆；序列分析；表达谱**

中图分类号：Q966　　文献标识码：A　　文章编号：0454-6296(2006)02-0172-07

Sequence analysis and expression of cathepsin D in the silkworm，*Bombyx mori*

YANG Yuan-Ping，LIU Chun，WANG Zi-Long，WANG Gen-Hong，XIA Qing-You* (The Key Sericultural Laboratory of Agricultural Ministry，College of Sericulture and Biotechnology，South West University，Chongqing 400716，China)

Abstract： Cathepsin D is lysosomal endoproteolytic enzyme which is thought to play key roles in the developmental and physiological process in organisms. The nucleic acid sequence of reported silkworm cathepdin D was used to tBLASTn search against the silkworm EST database. The ESTs with high score were clustered and assembled into a consensus sequence. Based on the consensus sequence，the aspartic proteases of *Bombyx mori* was cloned and identified，termed as *Bm*CtD (GenBank accession number：DQ010007). The cDNA was 1 543 bp at length with an open reading frame of 1 152 bp，and there were alternative splices in *Bm*CtD. The deduced amino acid sequence of the *Bm*CtD shared high similarity with that of cathepsin D in other species. The results of RT-PCR showed that the gene was expressed at all examined developmental stages and tissues of *B. mori*.

Key words： *Bombyx mori*；cathepsin D；clone；sequence analysis；expression

家蚕 *Bombyx mori* 是一种完全变态昆虫，经过卵、幼虫、蛹和成虫（蚕蛾）4 个形态和机能上完全不同的发育阶段而完成其生命周期。在昆虫的蜕皮与变态发育过程中，其形态和构造发生明显的阶段性变化，许多旧组织被系统地破坏，而一些未分化的细胞群发生分裂和生长，参与形成新的器官。在这一过程中同时伴随着丰富的组织蛋白酶的活动，它们在昆虫变态发育过程中对组织蛋白的水解起着重要的作用。

组织蛋白酶 D（cathepsin D，CtD）是天冬氨酸蛋白酶家族的重要成员，属于酸性溶酶体内切蛋白酶（Cho and Raikhel，1992），广泛存在于脊椎动物、真菌、反转录病毒和一些植物病毒中。通常情况下，组织蛋白酶 D 首先合成为被 6-磷酸甘露糖（m-6-p）修饰的糖蛋白，然后通过溶酶体靶受体经高尔基体转运至溶酶体，再被加工成稳定的活性形式，因此它主要以溶酶体酶存在发挥作用（Kornfeld，1990；Glickman and Kornfeld，1993）。组织蛋白酶 D 可以多

基金项目：国家重点基础研究发展规划项目（2005CB121000）；国家自然科学基金项目（30471313）；重庆市科委重点项目（20048571）

作者简介：杨远萍，女，1978 年 11 月生，硕士研究生，研究方向为生化与分子生物学，E-mail：y-y-p@163.com

* **通讯作者** Author for correspondence，E-mail：xiaqy@swau.cq.cn

收稿日期 Received：2005-05-23；**接受日期** Accepted：2005-10-17

种形式存在，分泌时为没有催化活性的酶原，通过自身降解为两条肽链组成的活性分子中间体，中间体进一步加工修饰为 1 条轻链和 1 条重链，为其具有活性的成熟形式（Huang *et al*.，1979；Hasilik and Neufeld，1980；Gieselmann *et al*.，1985；Erickson *et al*.，1989）。组织蛋白酶 D 在不同物种的各种组织中的功能有差异。在蚊子中，它特异地降解参与卵黄原蛋白生物合成的部分组织和器官（Cho and Raikhel，1992）。在血吸虫等寄生生物中，它参与降解寄主的血红蛋白（Maria *et al*.，2004）。在哺乳动物和脊椎动物中的研究发现，组织蛋白酶 D 与肿瘤的发生、发展密切相关，因而被当作是肿瘤侵袭性指标，倍受关注（张育超和吕永添，2004）。但是在昆虫的研究中，类似报道比较少。我们根据已公布的组织蛋白酶 D mRNA 序列、家蚕总的表达序列标签（expressed sequence tags，EST）和基因序列对家蚕组织蛋白酶 D 基因进行了结构分析，并调查了组织蛋白酶 D 在家蚕中的时空表达谱，为研究组织蛋白酶 D 在家蚕中的功能提供参考。

1　材料和方法

1.1　家蚕组织蛋白酶 D 基因的信息分析

从 Silkbase（http://silkworm.genomics.org.cn/index.jsp）中下载所有的 EST 序列，与本实验室的 EST 数据库构成家蚕的总 EST 数据库（http://www.ncbi.nlm.nih.gov/Genbank/index.html）。从 NCBI 下载家蚕组织蛋白酶 D 的 mRNA 序列（AY297160），对总 EST 数据库进行 BLASTN 检索，参数标准设置为 $E \leqslant e-30$，挑选出满足条件的序列，用 DNAstar 软件进行序列拼接延伸，经迭代检索到不能延伸为止。蛋白质 motif、信号肽和前导肽的预测在 EXPASY（http://au.expasy.org/）进行。家蚕组织蛋白酶 D 蛋白质序列与其他物种序列的比对应用 CLUSTALW。采用 Sim4 程序（http://gamay.univ-perp.fr/analyse_seq/sim4）进行基因的外显子、内含子边界分析。

1.2　实验材料

家蚕品种为华系二化性大造品种，按正常条件桑叶饲育。取幼虫到蛹期共 13 个时期的蚕个体、未受精卵、48 h 胚胎、戊期胚胎及 5 龄第 3 天的脂肪体、中肠、血液、丝腺、精巢和卵巢。液氮速冻后于 -80℃保存备用。

1.3　RNA 的提取和引物设计

按 Tripure 试剂（Roche）的说明书进行家蚕各时期和组织的 RNA 提取，按 AMV 反转录酶（Promega 公司）的操作说明反转录合成 cDNA 第 1 链，再进行 RT-PCR 扩增。引物序列为：F_1：5′-TTTACTACTTACTGCGTTGTT-3′；F_2：5′-ATGAAAACTGCGAGAACCCAC-3′；R_1：5′-GTATATTAATCTTAGTTCAGG-3′；R_2：5′-TAGCGTTGACATTGACAGAGT-3′。

1.4　家蚕组织蛋白酶 D cDNA 的克隆及序列测定

将 F_1R_1 引物组合对蚕蛹 cDNA 进行 PCR 扩增。反应条件：94℃预变性 5 min，然后按 94℃ 1 min，55℃ 1 min，72℃ 30 s，共 35 次循环，最后 72℃延伸 10 min。反应产物用 1.0% 琼脂糖凝胶电泳检测，用上海华舜胶回收试剂盒进行 DNA 回收，将回收的组织蛋白酶 D cDNA 片段与 pMD18-T 载体连接，进行 TA 克隆，对阳性克隆进行酶切鉴定，并用 ABI3700 测序仪测序验证。

1.5　表达谱分析

参照 TaKaRa RNA LA PCR™ Kit（AMV）Ver.1.1 说明书，用 F_1R_2 引物对家蚕 13 个时期、未受精卵、48 h 胚胎、戊期胚胎和 5 龄第 3 天各组织 cDNA 进行 RT-PCR 扩增，反应条件：94℃预变性 5 min，然后按 94℃ 1 min，55℃ 1 min，72℃ 30 s，共 35 次循环，最后 72℃延伸 10 min。反应产物用 2.0% 琼脂糖凝胶电泳检测。

2　结果

2.1　家蚕组织蛋白酶 D mRNA 的序列拼接及全长 cDNA 克隆

为了得到更为完整和准确的家蚕组织蛋白酶 D

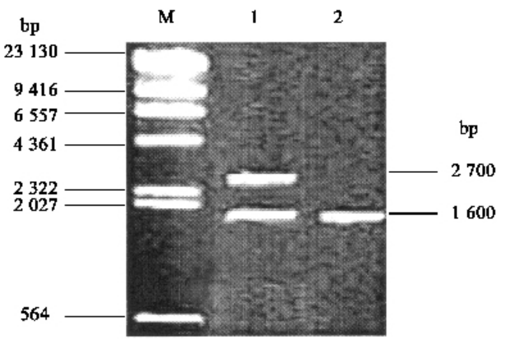


图 1　*Bm*CtD TA 克隆重组质粒酶切鉴定

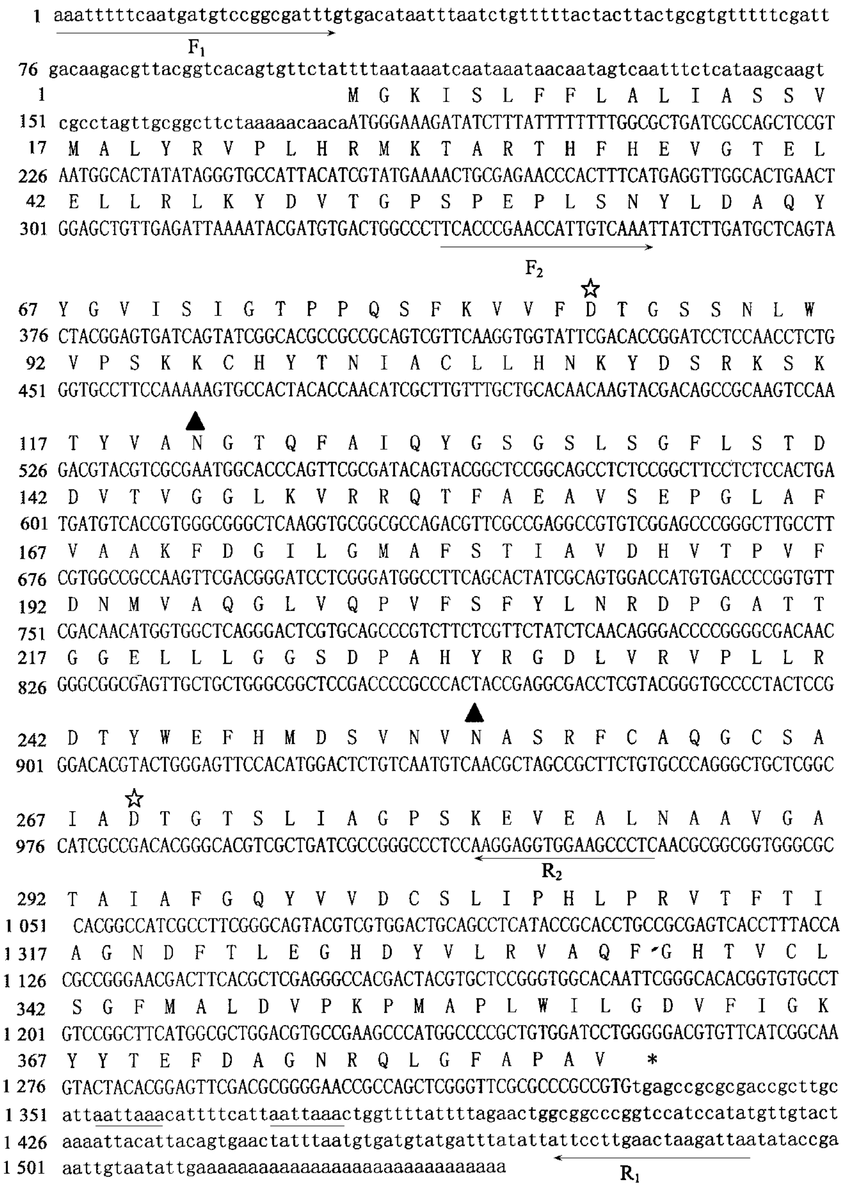
Fig. 1　The digested identification of recombinant plasmid of the *Bm*CtD TA clone

M：DNA 标准 Marker（λ-*Hin*d Ⅲ）；1. 重组质粒的酶切 Recombinant plasmid DNA digested with *Eco*R Ⅰ and *Hin*d Ⅲ；2. RT-PCR 产物 RT-PCR product.

基因序列，利用 tBLASTn **程序对家蚕总** EST **进行了序列比对，**E **值为** e－30，**将满足条件的序列以最小重叠** 100 nt **的等同值进行了序列拼接，得到一条长为** 1 543 bp **的全长** cDNA **序列，命名为** *Bm*CtD（GenBank **登录号**：DQ010007），**与目前所登录的家蚕组织蛋白酶** D **序列具有不同的多聚腺苷酸化位点；同时还发现来自未受精卵** cDNA **文库的** 3 **条组织蛋白酶** D EST **序列含有一段内含子序列，可能是组织蛋白酶** D mRNA **的一种异构体，命名为** *Bm*CtD Ⅰ。**提取家蚕蛹总** RNA，**进行** RT-PCR **扩增，经琼脂糖凝胶电泳分析出现单一约** 1 500 bp **的条带，与预计大小相符（图** 1），**用** pMD18-T **载体克隆后测序，结果与预测一致。**

2.2 *Bm*CtD 全长 cDNA 及预测蛋白序列分析

*Bm*CtD **基因的** cDNA **序列全长为** 1 543 bp，**开放阅读框**（ORF：179～1 330）**长为** 1 152 bp，**在** mRNA 1 354 bp**和** 1 371 bp **两处有** AATTAAA **加尾信号（图** 2）。**该基因编码** 384 **个氨基酸，**EXPASY **软件预测**



```
    1 aaatttttcaatgatgtccggcgatttgtgacataatttaatctgtttttactacttactgcgtgtttttcgatt
      --------------------------->
                  F1
   76 gacaagacgttacggtcacagtgttctattttaataaatcaataaataacaatagtcaatttctcataagcaagt
    1                              M  G  K  I  S  L  F  F  L  A  L  I  A  S  S  V
  151 cgcctagttgcggcttctaaaaacaacaATGGGAAAGATATCTTTATTTTTTTGGCGCTGATCGCCAGCTCCGT
   17  M  A  L  Y  R  V  P  L  H  R  M  K  T  A  R  T  H  F  H  E  V  G  T  E  L
  226 AATGGCACTATATAGGGTGCCATTACATCGTATGAAAACTGCGAGAACCCACTTTCATGAGGTTGGCACTGAACT
   42  E  L  L  R  L  K  Y  D  V  T  G  P  S  P  E  P  L  S  N  Y  L  D  A  Q  Y
  301 GGAGCTGTTGAGATTAAAATACGATGTGACTGGCCCTTCACCCGAACCATTGTCAAATTATCTTGATGCTCAGTA
                                                 ------------------------>
                                                            F2
                                                          ☆
   67  Y  G  V  I  S  I  G  T  P  P  Q  S  F  K  V  V  F  D  T  G  S  S  N  L  W
  376 CTACGGAGTGATCAGTATCGGCACGCCGCCGCAGTCGTTCAAGGTGGTATTCGACACCGGATCCTCCAACCTCTG
   92  V  P  S  K  K  C  H  Y  T  N  I  A  C  L  L  H  N  K  Y  D  S  R  K  S  K
  451 GGTGCCTTCCAAAAAGTGCCACTACACCAACATCGCTTGTTTGCTGCACAACAAGTACGACAGCCGCAAGTCCAA
                   ▲
  117  T  Y  V  A  N  G  T  Q  F  A  I  Q  Y  G  S  G  S  L  S  G  F  L  S  T  D
  526 GACGTACGTCGCGAATGGCACCCAGTTCGCGATACAGTACGGCTCCGGCAGCCTCTCCGGCTTCCTCTCCACTGA
  142  D  V  T  V  G  G  L  K  V  R  R  Q  T  F  A  E  A  V  S  E  P  G  L  A  F
  601 TGATGTCACCGTGGGCGGGCTCAAGGTGCGGCGCCAGACGTTCGCCGAGGCCGTGTCGGAGCCCGGGCTTGCCTT
  167  V  A  A  K  F  D  G  I  L  G  M  A  F  S  T  I  A  V  D  H  V  T  P  V  F
  676 CGTGGCCGCCAAGTTCGACGGGATCCTCGGGATGGCCTTCAGCACTATCGCAGTGGACCATGTGACCCCGGTGTT
  192  D  N  M  V  A  Q  G  L  V  Q  P  V  F  S  F  Y  L  N  R  D  P  G  A  T  T
  751 CGACAACATGGTGGCTCAGGGACTCGTGCAGCCCGTCTTCTCGTTCTATCTCAACAGGGACCCCGGGGCGACAAC
  217  G  G  E  L  L  L  G  G  S  D  P  A  H  Y  R  G  D  L  V  R  V  P  L  L  R
  826 GGGCGGCGAGTTGCTGCTGGGCGGCTCCGACCCCGCCCACTACCGAGGCGACCTCGTACGGGTGCCCCTACTCCG
                                              ▲
  242  D  T  Y  W  E  F  H  M  D  S  V  N  V  N  A  S  R  F  C  A  Q  G  C  S  A
  901 GGACACGTACTGGGAGTTCCACATGGACTCTGTCAATGTCAACGCTAGCCGCTTCTGTGCCCAGGGCTGCTCGGC
             ☆
  267  I  A  D  T  G  T  S  L  I  A  G  P  S  K  E  V  E  A  L  N  A  A  V  G  A
  976 CATCGCCGACACGGGCACGTCGCTGATCGCCGGGCCCTCCAAGGAGGTGGAAGCCCTCAACGCGGCGGTGGGCGC
                                                   <----------------
                                                          R2
  292  T  A  I  A  F  G  Q  Y  V  V  D  C  S  L  I  P  H  L  P  R  V  T  F  T  I
1 051 CACGGCCATCGCCTTCGGGCAGTACGTCGTGGACTGCAGCCTCATACCGCACCTGCCGCGAGTCACCTTTACCA
1 317  A  G  N  D  F  T  L  E  G  H  D  Y  V  L  R  V  A  Q  F  G  H  T  V  C  L
1 126 CGCCGGGAACGACTTCACGCTCGAGGGCCACGACTACGTGCTCCGGGTGGCACAATTCGGGCACACGGTGTGCCT
  342  S  G  F  M  A  L  D  V  P  K  P  M  A  P  L  W  I  L  G  D  V  F  I  G  K
1 201 GTCCGGCTTCATGGCGCTGGACGTGCCGAAGCCCATGGCCCCGCTGTGGATCCTGGGGGACGTGTTCATCGGCAA
  367  Y  Y  T  E  F  D  A  G  N  R  Q  L  G  F  A  P  A  V     *
1 276 GTACTACACGGAGTTCGACGCGGGGAACCGCCAGCTCGGGTTCGCGCCCGCCGTGtgagccgcgcgaccgcttgc
1 351 attaattaaacattttcattaattaaactggttttattttagaactggcggcccggtccatccatatgttgtact
         -------           -------
1 426 aaaattacattacagtgaactatttaatgtgatgtatgatttatattattccttgaactaagattaatataccga
                                                          <----------------------
1 501 aattgtaatattgaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa                         R1
```

图 2 *Bm*CtD **基因的** cDNA **序列及推导的氨基酸序列**

Fig. 2 The cDNA and deduced amino acid sequences of cathepsin D in *Bombyx mori*

F_1、F_2、R_1 **和** R_2 **分别表示上游引物和下游引物位置** The F_1, F_2, R_1, and R_2 position represented the upward and downward primer respectively；—— ：polyA **加尾信号** Polyadenylation signal was underlined；＊：**终止密码子** Stop codon was indicated by asterisk；☆：DTG motif **活性位点** DTG motif activity site；▲：**糖基化位点** The N-glycosylation site were indicated by dark triangle.

显示，从第1～18位氨基酸是信号肽，第19～53位是前导肽，成熟肽从第54～384位有331个氨基酸，分子量35.58 kD。结构域分析表明，*Bm*CtD属于典型的组织蛋白酶中的天冬氨酸蛋白酶家族。在第3和第8外显子中有与人、曼氏血吸虫组织蛋白酶D蛋白一样的DTG motif活性位点，两个糖基化位点分别位于第121和第255位氨基酸（图2）。通过NCBI的Blast（http://www.ncbi.nlm.nih.gov/BLAST/）程序进行在线蛋白质序列比对表明家蚕*Bm*CtD与NCBI登录的家蚕、冈比亚按蚊*Anopheles gambiae*、果蝇*Drosophila melanogaster*、桑天牛*Apriona germari*、小鼠*Mus musculus*、大鼠*Rattus norvegicus*、人*Homo sapiens*和曼氏血吸虫*Schistosoma mansoni*的蛋白质序列相似性分别是97%、76%、75%、73%、64%、65%、65%和64%，由此可以看出家蚕组织蛋白酶D与其他真核生物特别是昆虫的组织蛋白酶序列相似性较高，其序列相当保守。用*Bm*CtD基因的氨基酸序列与人（AAP36305.1）、小鼠（NP-599161.2）、大鼠（NP-034113.1）、鸡（Q05744）、牛（BAB21620.1）、按蚊（XP-307784.1）、埃及伊蚊（Q03168）、猪（KHPGD）、钩口线虫（AAO22152.1）、果蝇（AAF23824.1）、美洲板口线虫（CAC00543.1）、曼氏血吸虫（U60995.1）、斑马鱼（AJ278268.1）的组织蛋白酶D进行聚类分析，结果发现*Bm*CtD与果蝇、按蚊、埃及伊蚊和曼氏血吸虫亲缘关系较近，它们的DTG motif位置相近，且内含子位点大多保守，聚在一起，而与其他脊椎动物亲缘关系相对较远。但总体而言，蛋白酶家族序列保守性非常高（图3）。

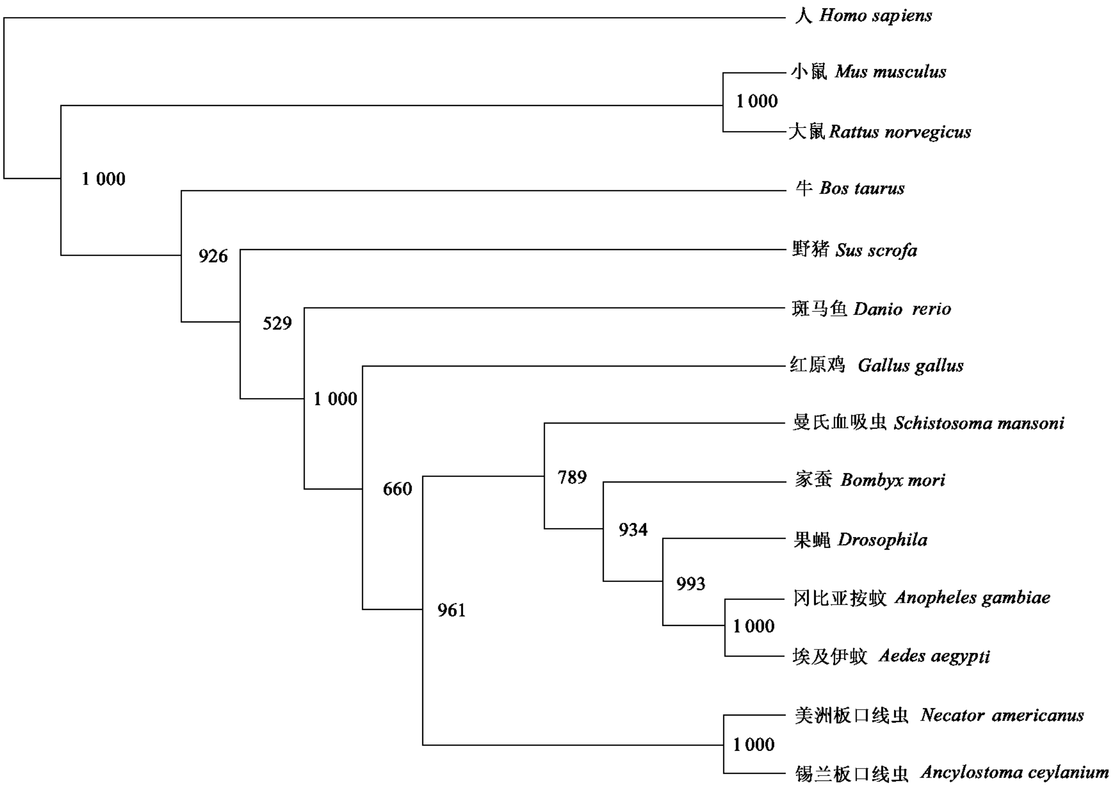


图3　*Bm*CtD与其他物种CtD的进化关系

Fig. 3　Phylogenetic relation of *Bombyx mori* aspartic protease with cathepsin D of other species

2.3　*Bm*CtD基因组结构分析

用Sim 4程序对*Bm*CtD基因的外显子/内含子结构进行分析，表明*Bm*CtD基因具有10个外显子（图4），且外显子/内含子边界均符合GT-AG规则（表1），*Bm*CtD Ⅰ有9个外显子，包括了*Bm*CtD的第1内含子（100 bp）在内（图4）。脊椎动物的天冬氨酸蛋白酶基因都有保守的9个外显子结构，人的CtD具有典型的脊椎动物基因结构的特征，活性位点DTG motif位于第3和第7外显子（Holm *et al*.，1984；Redecker *et al*.，1994）；曼氏血吸虫有7个外显子，DTG motif位于第3和第6外显子；而家蚕*Bm*CtD基因有10个外显子，DTG motif分别位于第3和第8外显子。这3个基因的内含子位点都很保守，有5个位点完全相同。在人、曼氏血吸虫中，信号肽都由第1外显子编码，前导肽则由第1外显子部分序列和第2外显子共同编码，家蚕中信号肽由第2外显子编码，前导肽由第2、3外显子编码（图5）。

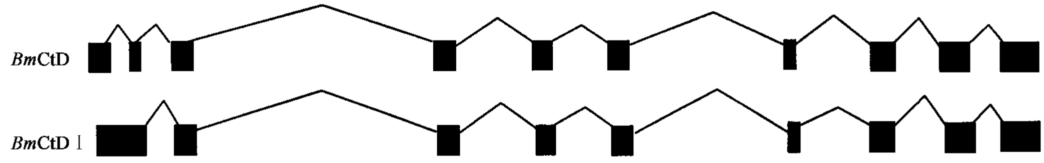
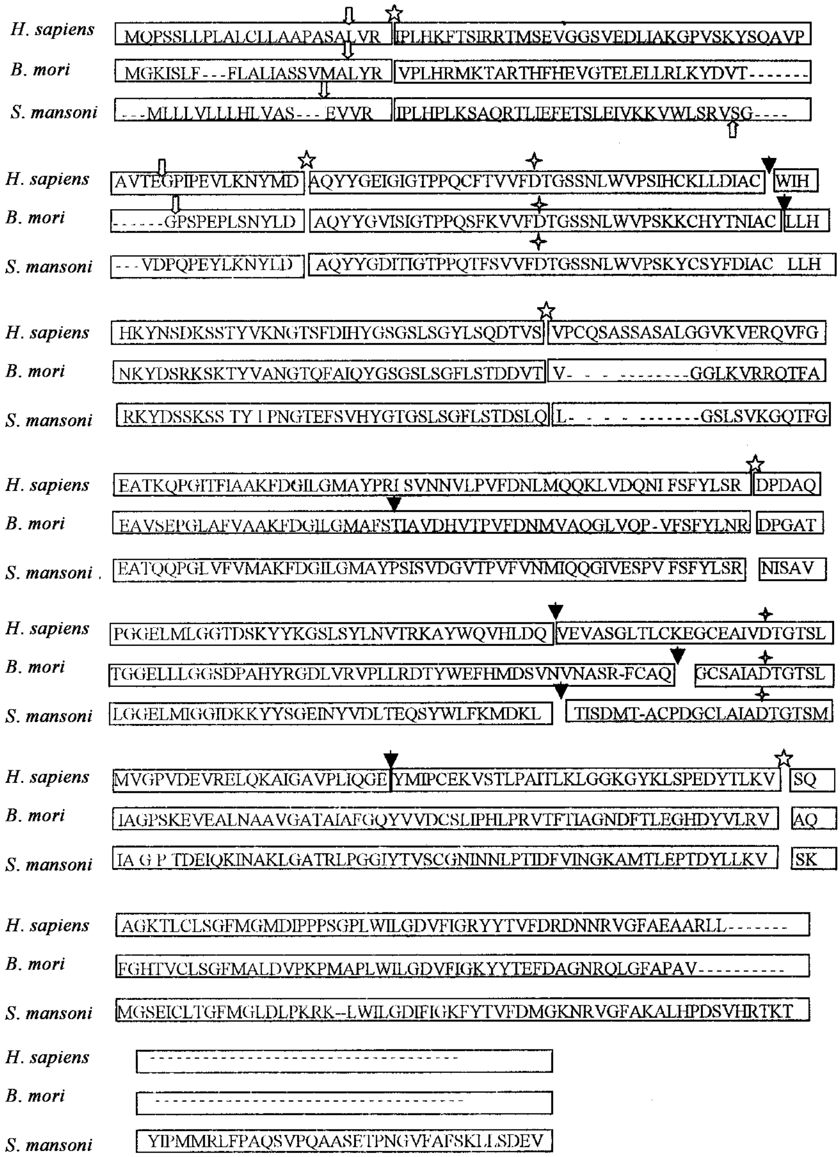


图 4 *Bm*CtD 基因结构

Fig. 4 The genome structure of *Bm*CtD gene

*Bm*CtD：10 个外显子 10 exons；*Bm*CtD Ⅰ：9 个外显子 9 exons.


H. sapiens
MQPSSLLPLALCLLAAPASALVR
IPLHKFTSIRRTMSEVGGSVEDLIAKGPVSKYSQAVP
B. mori
MGKISLF---FLALIASSVMALYR
VPLHRMKTARTHFHEVGTELELLRLKYDVT-------
S. mansoni
---MLLLVLLLHLVAS---EVVR
IPLHPLKSAQRTLIEFETSLEIVKKVWLSRVSG----
H. sapiens
AVTEGPIPEVLKNYMD
AQYYGEIGIGTPPQCFTVVFDTGSSNLWVPSIHCKLLDIAC
WIH
B. mori
------GPSPEPLSNYLD
AQYYGVISIGTPPQSFKVVFDTGSSNLWVPSKKCHYTNIAC
LLH
S. mansoni
---VDPQPEYLKNYLD
AQYYGDITIGTPPQTFSVVFDTGSSNLWVPSKYCSYFDIAC LLH
H. sapiens
HKYNSDKSSTYVKNGTSFDIHYGSGSLSGYLSQDTVS
VPCQSASSASALGGVKVERQVFG
B. mori
NKYDSRKSKTYVANGTQFAIQYGSGSLSGFLSTDDVT
V- - - -----------GGLKVRRQTFA
S. mansoni
RKYDSSKSS TY I PNGTEFSVHYGTGSLSGFLSTDSLQ
L- - - - -------GSLSVKGQTFG
H. sapiens
EATKQPGITFIAAKFDGILGMAYPRI SVNNVLPVFDNLMQQKLVDQNI FSFYLSR
DPDAQ
B. mori
EAVSEPGLAFVAAKFDGILGMAFSTIAVDHVTPVFDNMVAQGLVQP-VFSFYLNR
DPGAT
S. mansoni
EATQQPGLVFVMAKFDGILGMAYPSISVDGVTPVFVNMIQQGIVESPV FSFYLSR
NISAV
H. sapiens
PGGELMLGGTDSKYYKGSLSYLNVTRKAYWQVHLDQ
VEVASGLTLCKEGCEAIVDTGTSL
B. mori
TGGELLLGGSDPAHYRGDLVRVPLLRDTYWEFHMDSVNVNASR-FCAQ
GCSAIADTGTSL
S. mansoni
LGGELMIGGIDKKYYSGEINYVDLTEQSYWLFKMDKL
TISDMT-ACPDGCLAIADTGTSM
H. sapiens
MVGPVDEVRELQKAIGAVPLIQGE
YMIPCEKVSTLPAITLKLGGKGYKLSPEDYTLKV
SQ
B. mori
IAGPSKEVEALNAAVGATAIAFGQYVVDCSLIPHLPRVTFTIAGNDFTLEGHDYVLRV
AQ
S. mansoni
IA G P TDEIQKINAKLGATRLPGGIYTVSCGNINNLPTIDFVINGKAMTLEPTDYLLKV
SK
H. sapiens
AGKTLCLSGFMGMDIPPPSGPLWILGDVFIGRYYTVFDRDNNRVGFAEAARLL-------
B. mori
FGHTVCLSGFMALDVPKPMAPLWILGDVFIGKYYTEFDAGNRQLGFAPAV----------
S. mansoni
MGSEICLTGFMGLDLPKRK--LWILGDIFIGKFYTVFDMGKNRVGFAKALHPDSVHRTKT
H. sapiens

B. mori

S. mansoni
YIPMMRLFPAQSVPQAASETPNGVFAFSKLLSDEV

图 5 家蚕、曼氏血吸虫和人的组织蛋白酶 D 基因结构

Fig. 5 Comparison of the exon/intron structures of *Bombyx mori*, *Schistosoma mansoni* and human cathepsin D

✧：活性位点 DTG motif The position of the active site dyad of DTG motifs is indicated with asterisks；☆：内含子完全保守位点 White stars indicate where intron positions are conserved among the three enzymes；▼：有差异的内含子位点 Dark arrows indicate where introns occur in similar, though not exactly the same way；⇩：信号肽和前导肽剪切位点 The position of cleavage of signal peptides and the position of cleavage of the propeptides from mature enzymes were indicated with white arrows.

表 1　***Bm*CtD** 基因序列内含子/外显子位点

Table 1　Intron/exon sites of *Bm*CtD gene

外显子序号及长度 (bp) Exon number and size	cDNA	基因组 Scaff001627 genomic DNA	受位 3′ accepter	供位 5′ donor	内含子序号及长度 (bp) Intron number and size
1(147)	1 - 147	16 055 - 16 202	tatag/GTCGC	AAGCA/gtacg	1(100)
2(92)	148 - 240	16 302 - 16 394	tgcag/GGTGC	TATAG/gtaag	2(201)
3(126)	241 - 367	16 595 - 16 721	tacag/GCTCA	TTGAT/gtaag	3(2 098)
4(123)	368 - 491	18 819 - 18 942	tccag/TGCTG	TTGTT/gtcag	4(696)
5(208)	492 - 610	19 638 - 19 756	cgcag/GTGGG	TCACC/gtgag	5(104)
6(109)	611 - 720	20 156 - 20 265	tccag/TATCG	CGTAT/gtcgc	6(1 375)
7(86)	721 - 807	21 640 - 21 726	tccag/GGACC	GGTGG/gtacc	7(419)
8(156)	808 - 964	22 145 - 22 301	cccag/GGCTG	CCCAG/gttcg	8(314)
9(206)	965 - 1 171	22 615 - 22 821	gacag/GTGGCA	TCCGG/gtgag	9(149)
10(341)	1 172 - 1 513	22 970 - 23 308			

2.4　*Bm*CtD 基因表达谱分析

提取家蚕不同发育时期和 5 龄第 3 天各组织的 RNA，进行 RT-PCR，检测 *Bm*CtD 的表达情况。发育时期和组织分别包括 1 龄到 5 龄幼虫个体、吐丝前 3 天幼虫和第 1、3、5、7、9 天蛹个体，以及未受精卵、48 h胚胎、戊期胚胎以及 5 龄第 3 天的脂肪体、中肠、血液、丝腺、精巢和卵巢。结果表明，*Bm*CtD 在上述时期和组织中都有表达，而在未受精卵、48 h 胚胎和戊期胚胎中 *Bm*CtD 和 *Bm*CtD Ⅰ 两种 mRNA 形式都有表达(图 6 ~ 7)。

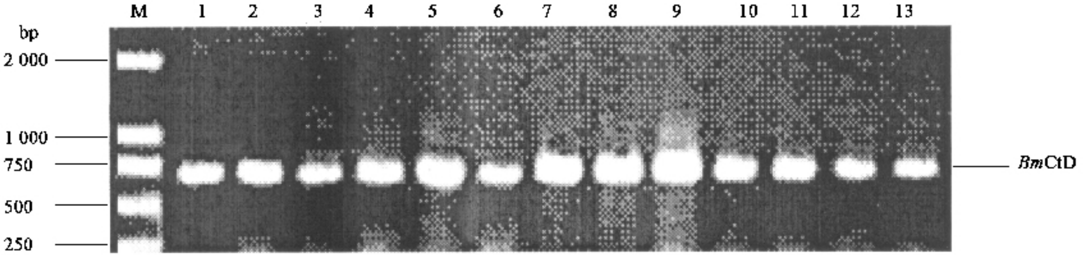


图 6　*Bm*CtD 基因在家蚕不同发育时期各组织 RT-PCR 扩增产物的电泳图

Fig. 6　Identification of the RT-PCR product of cDNA of *Bombyx mori* cathepsin D from at the different developmental stages

M：DL2000；1：1 龄幼虫 1st instar larva；2：2 龄幼虫 2nd instar larva；3：3 龄幼虫 3rd instar larva；4：4 龄幼虫 4th instar larva；5：5 龄幼虫 5th instar larva；6：上簇第 1 天 One day after wandering；7：上簇第 2 天 Two days after wandering；8：上簇第 3 天 Three days after wandering；9：第 1 天蛹 One day pupa；10：第 3 天蛹 3 days pupa；11：第 5 天蛹 5 days pupa；12：第 7 天蛹 7 days pupa；13：第 9 天蛹 9 days pupa.

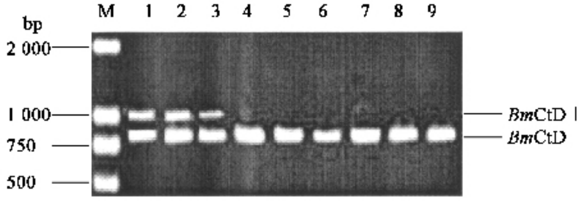


图 7　*Bm*CtD 基因在胚胎和 5 龄第 3 天各组织的表达情况

Fig. 7　Expression of the PCR product of cDNA of *Bombyx mori* cathepsin D from embryo and tissues of 5th instar larva

M：DL2000；1：未受精卵 Unfertilized egg；2：48 h 胚胎 48 h embryo；3：戊期胚胎 w-1 embryo；4：脂肪体 Fat body；5：中肠 Midgut；6：血液 Blood；7：丝腺 Silk gland；8：精巢 Testis；9：卵巢 Ovary.

3　讨论

昆虫体内的组织蛋白酶已有很多报道，其中，报道最多的是天冬氨酸蛋白酶家族的组织蛋白酶 D。组织蛋白酶 D 作为蛋白水解酶参与生物胚胎发育过程的功能已被许多学者所证实，特别对昆虫的发育变态十分重要。

NCBI 登录的家蚕 *Bm*CtD 全长 1 366 bp，编码 385 个氨基酸。本实验在此基础上得到长为 1 543 bp 的 cDNA，其中 ORF 长 1 152 bp，可编码 384 个氨基酸，推测二者在核酸和蛋白质水平上的细微差异，可能是品种差异导致的。基因结构分析显示，*Bm*CtD 基因具有 10 个外显子和 9 个内含子，此结构

与其他真核生物非常相似。9个内含子中有5个内含子位点与人和曼氏血吸虫一样，其余的也比较接近，而不同物种间的氨基酸序列相似性较高，由此说明该基因非常保守，推测其功能也有较高的保守性。

对家蚕 *Bm*CtD 的时空表达谱分析发现，*Bm*CtD 在本实验所调查的时期和组织中都有表达，这与它是一种广泛存在的天冬氨酸蛋白酶相符，其广谱性特征表明它是参与昆虫发育变态进程必不可少的蛋白水解酶。

在未受精卵、产卵后48 h胚胎和戊期胚胎内检测到大小分别为1.6 kb和1.5 kb两种类型的mRNA，而且，在未受精卵中两种选择性拼接形式的mRNA的表达量相当，而在发育的胚胎中其含内含子的mRNA的表达量明显低于不含内含子的mRNA正常形式，而在其他时期的组织中则只有1.5 kb形式的mRNA。在金头鲷鱼 *Sparus aurata* 的卵中也发现该基因存在选择性拼接，在上浮鱼卵中，CtD的mRNA接近1.9 kb，而在下沉鱼卵中，CtD mRNA的长度在1.3 kb左右，一般认为下沉鱼卵是质量好的能良好发育的（Carnevali *et al*.，1999）。对胚胎发育与组织蛋白酶D活性的研究认为组织蛋白酶D在卵黄原蛋白的降解过程中起着重要的作用，与卵的发育进程密切相关。*Bm*CtD 的这种选择性拼接不同的转录方式可能参与调控该蛋白的合成，甚至还可能与蚕卵的滞育性相关。由于 *Bm*CtD 基因的翻译起始位点位于第2外显子，在翻译起始位点的上游，是翻译识别因子及核糖体蛋白的结合位点，当mRNA含有100 bp的第1内含子时，该mRNA的翻译效率可能会受到影响甚至不发生翻译。当胚胎滞育时，该基因就主要以含内含子的这种形式进行转录；而胚胎解除滞育向前发育时，它则主要以不含内含子的这种形式进行转录，从而达到与胚胎发育需要卵黄原蛋白的降解提供营养相一致。目前在家蚕基因组中发现上百个基因有选择性拼接，这对信号传导、性别决定以及发育变态等生理过程的调控有着非常重要的意义（Zha *et al*.，2005）。*Bm*CtD 的选择性拼接是否与卵的滞育及卵发育质量的好坏有关，还需要进一步的验证。

参考文献（References）

Carnevali O，Centonze F，Brooks S，Marota I，Sunpter JP，1999. Molecular cloning and expression of ovarian cathepsin D in seabream，*Sparus aurata*. *Biology of Reproduction*，61：785－791.

Cho WL，Raikhel AS，1992. Cloning of cDNA for mosquito lysosomal aspartic protease：Sequence analysis of an insect lysosomal enzyme similar to cathepsins D and E. *J. Biol. Chem.*，267(30)：21 823－21 829.

Erickson AH，1989. Biosynthesis of lysosomal endopeptidases. *J. Cell Biochem.*，40：31－41.

Gieselmann V，Hasilik A，Von Figura K，1985. Processing of human cathepsin D in lysosomes *in vitro*. *J. Biol. Chem.*，260：3 215－3 220.

Glickman JN，Kornfeld S，1993. Mannose 6-phosphate-independent targeting of lysosomal enzymes in I-cell disease B lymphoblasts. *J. Cell Biol.*，123：99－108.

Hasilik A，Neufeld EF，1980. Biosynthesis of lysosomal enzymes in fibroblasts synthesis as precursors of higher molecular weight. *J. Biol. Chem.*，255：4 937－4 945.

Holm I，Ollo R，Panthier JJ，Rougeon F，1984. Evolution of aspartyl proteases by gene duplication：the mouse renin gene is organized by two homologous clusters of four exons. *Embo. J.*，3：557－562.

Huang JS，Huang SS，Tang J，1979. Cathepsin D isoenzymes from porcine spleens：large scale purification andpolypeptide chain arrangements. *J. Biol. Chem.*，254：11 405－11 417.

Kornfeld S，1990. Lysosomal enzyme targeting. *Biochem. Soc. Trans.*，18：367－374.

Morales ME，Kalinna BH，Heyers O，Mann VH，Schulmeister A，Copeland CS，Loukas A，Brindley PJ，2004. Genomic organization of the *Schistosoma mansoni* aspartic protease gene，a platyhelminth orthologue of mammalian lysosomal cathepsin D. *Gene*，338(1)：99－109.

Redecker B，Heckendorf B，Grosch HW，Mersmann G，Hasilik A，1994. Molecular organization of the human cathepsin D gene. *DNA Cell Biol.*，10：423－431.

Zha XF，Xia QY，Zhao P，Li J，Duan J，Wang ZL，Qian JF，Xiang ZH，2005. Detection and analysis of alternative splicing in the silkworm by aligning expressed tags with the genomic sequence. *Insect Mol. Biol.*，14(2)：113－119.

Zhang YC，Lu YT，2004. Relationship between expression of cathepsin-D and nm23-H1 and infiltration，metastasis of colorectal cancer. *Lingnan Modern Clinics in Surgery*，4(1)：20－22. [**张育超，吕永添**，2004. **组织蛋白酶**-D**和** nm23-H1 **蛋白表达与结直肠癌浸润、转移的关系. 岭南现代临床外科**，4(1)：20－22]

（责任编辑：黄玲巧）